Determining distribution parameters from quantiles

John D. Cook
Department of Biostatistics
The University of Texas M. D. Anderson Cancer Center
P. O. Box 301402 Unit 1409
Houston, TX 77230-1402 USA
cook@mdanderson.org

January 27, 2010

Abstract

Bayesian statistics often requires eliciting prior probabilities from subject matter experts who are unfamiliar with statistics. While most people an intuitive understanding of the mean of a probability distribution, fewer people understand variance as well, particularly in the context of asymmetric distributions. Prior beliefs may be more accurately captured by asking experts for quantiles rather than for means and variances.

This note will explain how to solve for parameters so that common distributions satisfy two quantile conditions. We present algorithms for computing these parameters and point to corresponding software.

The distributions discussed are normal, log normal, Cauchy, Weibull, gamma, and inverse gamma. The method given for the normal and Cauchy distributions applies more generally to any location-scale family.

1 Problem statement

Let X be a random variable from a two-parameter family. Given probabilities p_1 and p_2 we elicit values x_1 and x_2 such that

$$P(X < x_1) = p_1$$

and

$$P(X < x_2) = p_2.$$

For example, we may ask for the 10th and 90th percentiles. For consistency we require

$$(x_1 - x_2)(p_1 - p_2) > 0.$$

This is the only restriction. There is no mathematical requirement for p_1 and p_2 to be widely separated, though in practice they usually will be.

2 Location-scale families

Suppose we want to find the mean and standard deviation of a normal distribution that has specified quantiles. The derivation below shows how to compute the distribution parameters. The only property of the normal distribution we use is that it is a location-scale family. The same derivation shows how to find the location and scale of any location-scale distribution.

We want to find μ and σ such that a normal random variable with mean (location) μ and standard deviation (scale) σ satisfies

$$P(X < x_1) = p_1$$

and

$$P(X < x_2) = p_2$$

where $x_1 < x_2$ and $p_1 < p_2$.

The random variable X has the same distribution as $\sigma Z + \mu$ where Z is a normal random variable with mean 0 and standard deviation 1. Let Φ be the CDF of Z, i.e. $\Phi(x) = P(Z < x)$. Then the equations

$$P(X < x_i) = p_i$$

are equivalent to

$$P(\sigma Z + \mu < x_i) = p_i$$

and so

$$P\left(Z < \frac{x_i - \mu}{\sigma}\right) = \Phi\left(\frac{x_i - \mu}{\sigma}\right) = p_i.$$

This yields linear equations for the parameters

$$\Phi^{-1}(p_i)\sigma + \mu = x_i$$

with the solution

$$\sigma = \frac{x_2 - x_1}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)}$$

and

$$\mu = \frac{x_1\Phi^{-1}(p_2) - x_2\Phi^{-1}(p_1)}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)}.$$

A log-normal distribution can easily be determined by two quantiles by reducing the problem to that of finding the parameters for a normal distribution.

If $X \sim$ log-normal(μ, σ), then $Y = \log X \sim$ normal(μ, σ). To find μ and σ so that $P(X < x_i) = p_i$, solve for parameters so that $P(Y < \log x_i) = p_i$.

Note that the argument above could be used to find the location and scale parameters for any location-scale distribution. For example, the procedure could be used to find parameters of a Cauchy distribution. If Z is the distribution family representative with location 0 and scale 1 and $F(x)$ is its CDF, then the scale parameter

$$\sigma = \frac{x_2 - x_1}{F^{-1}(p_2) - F^{-1}(p_1)}$$

and the location parameter

$$\mu = \frac{x_1 F^{-1}(p_2) - x_2 F^{-1}(p_1)}{F^{-1}(p_2) - F^{-1}(p_1)}$$

guarantee that $X = \sigma Z + \mu$ satisfies

$$F(x_1) = p_1$$

and

$$F(x_2) = p_2.$$

3 Weibull distribution

Next we consider the problem of determining the parameters of a Weibull distribution from two specified quantiles.

The Weibull distribution with shape γ and scale β has CDF

$$F(x; \gamma, \beta) = 1 - \exp\left(-\left(\frac{x}{\beta}\right)^{\gamma}\right).$$

The CDF can be inverted easily:

$$F^{-1}(p; \gamma, \beta) = \beta(-\log(1-p))^{1/\gamma}.$$

If

$$F(x_1; \gamma, \beta) = p_1$$

then

$$F(x_1/\beta; \gamma, 1) = p_1.$$

Thus for every $\gamma > 0$

$$\beta = \beta(\gamma) = \frac{x_1}{F^{-1}(p_1; \gamma, 1)} = \frac{x_1}{(-\log(1-p_1))^{1/\gamma}}$$

satisfies the first quantile equation $F(x_1; \gamma, \beta) = p_1$. Next we show how to select γ to also satisfy the second quantile equation.

If $F(x_i; \gamma, \beta(\gamma)) = p_i$ for $i = 1, 2$ then $x_i = F^{-1}(p_i; \gamma, 1)$ and

$$\frac{x_2}{x_1} = \frac{F^{-1}(p_2; \gamma, 1)}{F^{-1}(p_1; \gamma, 1)} = \left(\frac{\log(1 - p_2)}{\log(1 - p_1)} \right)^{1/\gamma}.$$

It follows that

$$\gamma = \frac{\log(-\log(1 - p_2)) - \log(-\log(1 - p_1))}{\log(x_2) - \log(x_1)}.$$

4 Gamma distribution

We can determine the parameters for a gamma distribution in a manner similar to that used for the Weibull distribution. However, the CDF and inverse CDF of a gamma distribution do not have an elementary closed form and so the proof is less direct.

Let $F(x; \alpha, \beta)$ be the CDF of a gamma distribution with shape α and scale β. That is,

$$F(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^x t^{\alpha-1} \exp(-t/\beta)\, dt.$$

We will show there is a unique solution to the equations

$$F(x_1; \alpha, \beta) = p_1$$

and

$$F(x_2; \alpha, \beta) = p_2$$

provided $0 < x_1 < x_2$ and $0 < p_1 < p_2 < 1$.

Note that the gamma is a scale family and so

$$F(x; \alpha, \beta) = F(x/\beta; \alpha, 1).$$

Thus for any $\alpha > 0$,

$$\beta = \frac{x_1}{F^{-1}(p_1; \alpha, 1)}$$

is the unique β satisfying

$$F(x_1; \alpha, \beta) = p_1.$$

This says that set of parameters satisfying one quantile constraint is a curve in the first quadrant. This curve implicitly defines β as a function of α. The question now is whether the curve corresponding to two different constraints must intersect.

Assume $x_1 > 0$ and $0 < p_1 < p_2 < 1$ are given. We consider the range of x_2 values such that the two quantile constraints can be satisfied. We will show that this range is (x_1, ∞).

If we can find $\alpha > 0$ such that

$$\beta = \frac{x_1}{F^{-1}(p_1; \alpha, 1)} = \frac{x_2}{F^{-1}(p_2; \alpha, 1)}$$

then (α, β) satisfies both constraints $F(x_i; \alpha, \beta) = p_i$. For every α, we can obtain equality above when

$$x_2 = x_1 \frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_2; \alpha, 1)}.$$

The question now is reduced to determining the range of

$$\frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)}.$$

First we show

$$\lim_{\alpha \to \infty} \frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)} = 1.$$

As $\alpha \to \infty$, the CDF of a gamma(α, 1) random variable approaches the CDF of a normal(α, $\sqrt{\alpha}$) random variable. Thus for large α

$$\frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)} \sim \frac{\sqrt{\alpha}\Phi^{-1}(p_2) + \alpha}{\sqrt{\alpha}\Phi^{-1}(p_1) + \alpha} \to 1.$$

Next, we show

$$\lim_{\alpha \to 0} \frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)} = \infty.$$

For small values of α,

$$1/\Gamma(\alpha) \sim \alpha$$

and

$$\int_0^x t^{\alpha-1} \exp(-t)\, dt \sim \frac{x^\alpha}{\alpha}.$$

This says that as $\alpha \to 0$,

$$F(x; \alpha, 1) \sim x^\alpha$$

and so

$$F^{-1}(p_i; \alpha, 1) \sim p_i^{1/\alpha}.$$

Therefore since $p_2 > p_1$ we have

$$\frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)} \sim \left(\frac{p_2}{p_1}\right)^{1/\alpha} \to \infty$$

as $\alpha \to 0$.

This says that for fixed $0 < p_1 < p_2 < 1$ and $0 < x_1 < x_2$, we can always find a value of α such that

$$\frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)} = \frac{x_2}{x_1} \tag{1}$$

and hence we can satisfy both quantile conditions. To compute α we can solve the equation (1), say using a bisection method.

To prove that that the solution produced above is unique it suffices show that that

$$f(\alpha) = \frac{F^{-1}(p_2; \alpha, 1)}{F^{-1}(p_1; \alpha, 1)}$$

is monotone decreasing. We have shown that $f(\alpha)$ is monotone decreasing for sufficiently small and sufficiently large values of α via asymptotic arguments. A complete proof would show that $f(\alpha)$ is monotone for all α.

Note that we could solve for the parameters of an inverse gamma random variable X by reducing the problem to finding parameters for the gamma random variable $1/X$. We have $P(X < x_i) = p_i$ if and only if $P(Y < 1/x_i) = 1 - p_i$.

5 Software

ParameterSolver is a Windows application that provides a convenient user interface for determining distribution parameters to satisfy two quantile conditions. The software also determines distribution parameters given a mean and variance. The software is available from the M. D. Anderson Cancer Center Biostatistics software download site.

`http://biostatistics.mdanderson.org/SoftwareDownload/`

This software does not implement the algorithms developed here. Instead it uses optimization to find the parameters that minimize

$$(F(x_1) - p_1)^2 + (F(x_2) - p_2)^2 .$$

I did not realize at the time I worked on the software that the quantile constraints could be satisfied exactly or that they could so easily be calculated directly for several distributions. The fact that the software provided parameters that exactly satisfy the quantile constraints prompted this more theoretical note.

The software will solve for beta distribution parameters satisfying two quantile conditions. It appears that these conditions can be satisfied exactly, though I have not proven this.